

А.В. ИВАШКО, канд. техн. наук, проф. НТУ «ХПИ»,
А.И. ПОТАПЕНКО, аспирант НТУ «ХПИ»

АЛГОРИТМЫ ВЫДЕЛЕНИЯ ОБЪЕКТОВ НА ИЗОБРАЖЕНИИ

У статті розглянути алгоритми виділення об'єктів на статичних зображеннях. Проведено порівняльний аналіз методів виділення об'єктів з навчанням по прецедентах. На основі експерименту висунуті припущення щодо оптимального алгоритму пошуку об'єктів.

В статье рассмотреть алгоритмы выделения объектов на статических изображениях. Проведен сравнительный анализ методов выделения объектов с учебой по прецедентам. На основе эксперимента выдвинутые предположения относительно оптимального алгоритма поиска объектов.

The algorithms for object detection on static images are considered in the article. Comparative analysis of the machine-learning object detection methods is carried out. Assumption about the optimal detection algorithm is made on the basis of experimental investigations.

Задачей систем компьютерного зрения является анализ снимаемой с датчиков видеoinформации и получение символьной интерпретации, определяющей характер изображений, присутствующих в сцене, и выявление сравнительного пространственного взаимоположения объектов.

Одной из важных частных задач машинного зрения является задача распознавания объектов на изображениях, то есть определения, является ли предъявленное изображение отображением интересующего нас объекта. Распознавание объектов является, в общем случае, некорректно поставленной задачей (легко представить два объекта разной природы, одинаково выглядящих на изображении). Люди распознают множество объектов на изображениях с незначительными усилиями, несмотря на то, что вид объектов может изменяться в зависимости от точки наблюдения, удаления, угла зрения, освещения и т.д. Объекты можно распознать, даже если они частично скрыты из вида. Но для систем компьютерного зрения такая задача до сих пор является не до конца решенной.

Рассмотрим различные таксономии для методов классификации в распознании образов. К настоящему времени ни одна из них не имеет бесспорного преимущества по сравнению с конкурентами. Наиболее принципиальная дихотомия проходит между статистическими и синтаксическими классификаторами.

Статистические методы принятия решения могут быть далее разделены различными путями, в зависимости от тех свойств, которым мы хотим уделить особое внимание. Одна из часто используемых дихотомий – противопоставление параметрических и непараметрических методов. В параметрических методах особая функциональная форма подразумевается как функция плотностей вектора свойств, в то время как непараметрические методы ссы-

лаются непосредственно на доступные данные. Используются также полупараметрические методы, которые пытаются объединить лучшие качества обоих подходов, используя ограниченное число адаптируемых параметров в зависимости от внутренней сложности данных.

Часто используется также деление на нейронные и классические статистические методы. Это имеет смысл, только, если мы хотим рассматривать эти методы как полностью не связанные конкурирующие подходы. Другая крайность - это рассмотрение нейронных методов только как особый итеративный подход для получения классических результатов традиционных статистических методов.

Нейронные методы могут быть дополнительно охарактеризованы по процессу обучения: алгоритмы “с учителем” нуждаются в заранее классифицированных данных для обучения, в то время как алгоритмы “без учителя” могут также использовать некатегоризированные данные. Ввиду природы процесса классификации, в большинстве случаев применимы только алгоритмы “с учителем”.

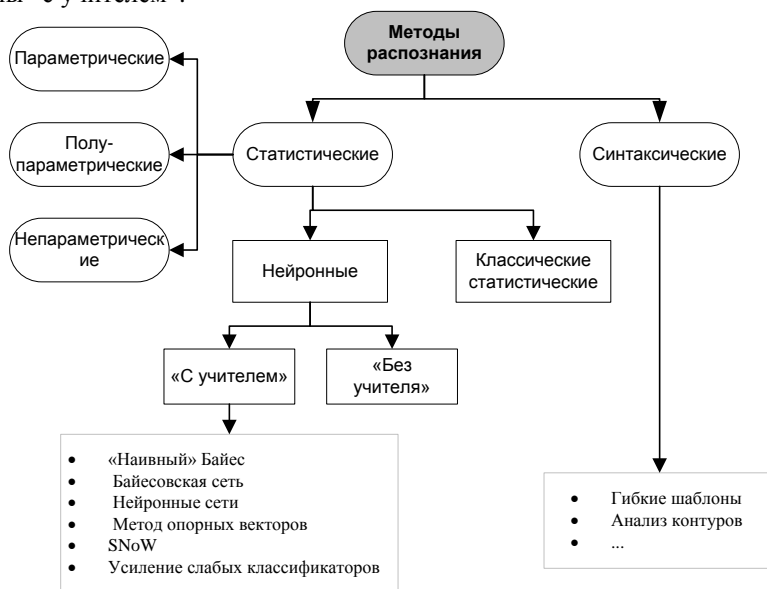


Рис. Методы распознавания изображений

Может возникнуть вопрос, почему мы предпочитаем именно статистические методы всем остальным. На самом деле, как с прагматичной, так и с теоретической точек зрения, преимущества статистических методов весьма убедительны:

- Сигналы с сенсоров и их характерные свойства проявляют вероятностный характер из-за шумов датчиков, изменяющихся условий освещения и ошибок алгоритмов предварительной обработки кадра;

- В общем случае, системы выделения объектов должны использовать всю имеющиеся источники информации, в том числе априорные знания и эмпирические данные. Требуется унифицированная математическая формулировка объединяющая все компоненты. В статистических алгоритмах, априорные знания кодируются в априорных вероятностных характеристиках объектов или их множественного появления в более сложных сценах (вероятностное моделирование сцены).

Стохастические свойства эмпирических данных суммируются в вероятностных коэффициентах характеризующих классы объектов. Согласно Байесовской теореме апостериорные данные объединяют в себе априорные данные и вероятностные характеристики классов в общем случае с помощью простого умножения;

- Еще одним фундаментальным аргументом в пользу статистических методов является то, что теория принятия решений гарантирует оптимальность Байесовских классификаторов, которые максимизируют апостериорные вероятности. Все классификаторы стремятся обеспечить модели для расчета (теоретически) оптимальных решений. Это соответствует аппроксимации апостериорной информации полученной из заданных данных для обучения. Использование статистических методов является прямым и наиболее естественным решением;

- Структуры алгоритмов обучения могут использовать обширные ресурсы статистики и теории статистического обучения. Если, к примеру, оценки апостериорных распределений основаны на параметрических функциях плотностей вероятности, то могут использоваться стандартные методики оценки параметров.

В дополнение к этим общим и фундаментальным преимуществам, вероятностный подход предлагает некоторые полезные инструменты для упрощения и увеличения вычислительной разрешимости; объединение независимых предположений о наблюдаемых характерных чертах ведет к компромиссным решениям и прокладывает путь к разрешению противоречия между вычислительной эффективностью и подробностью моделей, которая должна обеспечить достаточную дискриминантную способность. Избавляясь от случайных переменных путем их интеграции, мы снижаем сложность, получаем возможность использования вероятностных моделей при неполных входных данных, и обеспечиваем методику для задания схем распознавания с иерархическими моделями.

Мы получим оптимальный классификатор, если сможем охарактеризовать статистические параметры объектов появляющихся в данных с сенсоров. Но обычно это весьма нетривиальная задача, которая сводится к фундаментальной проблеме определения и вычисления апостериорных данных на основе эмпирических данных. Без соответствующих вероятностных моделей и аккуратной аппроксимации апостериорных данных невозможно реализовать оптимальную систему выделения объектов.

Распознавание "снизу-вверх" использует инвариантные свойства изображений объектов, опираясь на то предположение, что если человек может без усилий распознать тот или иной объект на изображении независимо от его ориентации, условий освещения и индивидуальных особенностей, то должны существовать некоторые признаки присутствия объектов этого типа на изображении, инвариантные относительно условий съемки. Алгоритм работы методов распознавания "снизу-вверх" может быть кратко описан следующим образом:

1 Обнаружение элементов и особенностей, которые характерны для изображения объекта;

2 Анализ обнаруженных особенностей, вынесение решения о количестве и расположении объектов;

На основе обнаруженных характерных свойств формируется модель класса объектов. Модель строится автоматически по заранее собранному набору прецедентов - изображений, для которых известно, являются ли они изображениями объекта или нет. Наблюдением в данном случае, является некоторый «вектор признаков», полученный из исходного изображения некоторым преобразованием, отображающим изображения в пространство действительных векторов. Гипотеза, подлежащая проверке, - принадлежность изображения к классу изображений искомого объекта.

Таким образом, система распадается на два модуля: модуль преобразования изображения в вектор признаков и модуль классификации. Задачей модуля преобразования является наиболее полное и информативное представление изображения в виде числового вектора. Задачей модуля классификации является проверка гипотезы принадлежности изображения классу изображений объекта на основании наблюдения, которым является вектор признаков.

Модуль преобразования и модуль классификации тесно связаны. Главная цель модуля преобразования – представить изображение в форме наиболее удобной для модуля классификации. Основные требования, предъявляемые к модулю преобразования: скорость, наиболее полное и информативное представление данных, масштабируемость. Также является желательной инвариантность модуля преобразования к деформациям и пространственным искажениям объекта.

Проанализируем существующие методы классификации с обучением по прецедентам, применимые к задаче распознавания объектов на изображениях. Данные методы выбраны к рассмотрению, ввиду их общности и многообещающих результатов в реальных системах.

Опишем формальную постановку задачи для модуля классификации:

Вектор признаков $x \in X \subseteq R^n$ - является описанием объекта, которое представляется модулем преобразования.

Классом будем называть некоторое подмножество $K_y = \{x \in X | y^*(x) = y\}$ множества X . Пусть $y \in Y \subset Z$ - множество маркеров классов, будем рас-

сма́тривать случай бинарной классификации, то есть $Y=\{-1;+1\}$. Применительно к задаче распознавания объектов подразумевается два варианта – изображение есть изображение искомого объекта или какого-либо другого.

$X \xrightarrow{y^*} Y$ - отображение, определённое для всех $x \in X$, задающее разбиение X на подмножества K_y , которые могут пересекаться. Причина возможности пересечения классов, заключается в том, что вектор признаков есть всего лишь описание объекта, а не сам объект: описание может быть неточным и для двух разных объектов описания могут совпадать.

Обучающей выборкой называется набор пар (набор прецедентов) $T=(x_1,y_1), \dots, (x_b,y_b)$ для которых $y^*(x_i) = y_i, i = \overline{1, l}$ то есть это известная информация об отображении $X \xrightarrow{y^*} Y$.

Для применения алгоритмов классификации и распознавания образов принимается следующая гипотеза:

Гипотеза: Множество $X \times Y$ является вероятностным пространством с вероятностной мерой P . Прецеденты $(x_1,y_1), \dots, (x_b,y_b)$ появляются случайно и независимо в соответствии с распределением P .

Задача классификации – построить функцию $F(x)$, классификатор, приближающую отображение y^* , основываясь на обучающей выборке $(x_1,y_1), \dots, (x_b,y_b)$.

Эмпирическим риском будем называть $P(F(x) \neq y | (x, y) \in T)$ то есть вероятность неверной работы классификатора для векторов признаков из обучающей выборки.

Общим риском будем называть $P(F(x) \neq y | x \in X)$, то есть вероятность того, что классификатор ошибётся на данных, не входивших в обучающую выборку. Основной целью при построении классификатора является минимизация именно общего риска. Так как напрямую вычислить величину общего риска невозможно, для проверки качества классификатора используется оценка ошибки на *контрольной выборке*, которая состоит из прецедентов, не входящих в обучающую выборку. Говорят, что классификатор обладает хорошей *обобщающей способностью*, если при обучении классификатор эффективно уменьшает общий риск, оценённый на контрольной выборке.

Алгоритм построения склонен к *переобучению*, если при минимизации эмпирического риска, общий риск начинает возрастать. Этот феномен связан с тем, что классификатор начинает обобщать признаки свойственные не данным в целом, конкретно прецедентам из обучающей выборки.

Байесовские методы классификации

Принцип максимума апостериорной вероятности основывается на трёх гипотезах:

1 Множество $X \times Y$ является вероятностным пространством с вероятностной мерой P . Прецеденты $(x_1,y_1), \dots, (x_b,y_b)$ появляются случайно и независимо в соответствии с распределением P ;

2 Известны плотности распределения классов $p_y(x) = p(x|K_y), y \in Y$, называемые функциями правдоподобия;

3 Известны вероятности появления объектов каждого из классов $P_y = P(K_y), y \in Y$, называемые априорными вероятностями.

При принятии данных гипотез решающее правило записывается в аналитическом виде:

$$F(x) = \arg \max_{y \in Y} P(K_y|x) = \arg \max_{y \in Y} p_y(x)P_y \quad (1)$$

Такой выбор решающего правила является оптимальным с точки зрения минимизации общего риска. Разделяющая поверхность в данном случае не имеет какой-либо чёткой геометрической структуры.

Существуют системы обнаружения объектов на изображениях основанные на «наивном» Байесовском методе. Данный метод основывается на построении эмпирической плотности распределения вероятностей классов по обучающей выборке в предположении о независимости компонентов вектора признаков.

Проблема слабой обусловленности решается за счет выбора компактного представления вектора признаков. Практические результаты следующие:

- 1 В задаче распознавания лиц примерно 91% верных обнаружений и порядка 0,2% ложных срабатываний;
- 2 Алгоритм построения «наивного» Байесовского классификатора склонен к переобучению и чувствителен к шуму;
- 3 Скорость работы самого классификатора крайне высока, основное время может занимать вычисление вектора признаков.

Классические нейронные сети

Под термином нейронные сети скрывается целый класс алгоритмов. Основная идея, лежащая в основе нейронных сетей - это последовательное преобразование сигнала, параллельно работающими элементарными функциональными элементами, нейронами. Основной принцип настройки нейронной сети заключается в применении оптимизационных методов к минимизации среднеквадратичной ошибки, как следствие – склонность к переобучению. Главное преимущество нейронных сетей – гибкость. Возможность эвристических модификаций нейросетей почти безграничны. Нейросети относятся к классическим методам машинного зрения, хотя в последнее время стали терять популярность.

Системы распознавания объектов на изображениях, основанные на нейронных сетях, используют иерархическую архитектуру. Вначале вектор признаков обрабатывается грубой сетью с высоким уровнем ошибок второго рода, далее, если вектор не был классифицирован как не объект, решение корректируется более точной и более медленной сетью. Такой подход даёт вполне приемлемые результаты:

1 92% правильных обнаружений при 1,3% ошибок второго рода для задачи распознавания лиц;

2 Нейронные сети в целом склонны к переобучению, хотя и существуют некоторые методы, которые в частном случае могут решить эту проблему;

3 Устойчивость к шуму сильно зависит от конкретной архитектуры сети. В общем случае, нейросеть чувствительна к шуму;

4 Вычислительная сложность квадратично зависит от числа нейронов в скрытом слое.

Метод опорных векторов

Классический алгоритм заключается в построении линейной разделяющей поверхности (гиперплоскости), равноудалённой от выпуклых оболочек классов, выпуклая оболочка строится по прецедентам. Утверждается, что такая разделяющая гиперплоскость будет оптимальна, с точки зрения общего риска, относительно любых других возможных гиперплоскостей. Метод опорных векторов был успешно применён для задачи распознавания объектов на изображениях. Подход даёт следующие результаты:

1 При порядка 72% верных обнаружений порядка 0,6% ложных срабатываний, для задачи распознавания лиц.

2 Очень высокая устойчивость к переобучению

3 Чувствительность к шуму может регулироваться, за счет уменьшения точности. Обычно стараются найти компромисс между устойчивостью и точностью.

4 В системах распознавания объектов на изображениях метод даёт ускорение в несколько раз, по сравнению с нейронными сетями.

Разреженная просеивающая сеть

SNoW (Sparse network of Winnows) – особый вид нейронной сети. Вектор признаков предполагается бинарным. Сеть состоит из двух (по числу возможных классов) линейных нейронов, связанных с компонентами вектора признаков. Классификация проходит по принципу победитель забирает всё.

SNoW считается достаточно эффективным методом для решения задачи обнаружения объектов на изображениях. SNoW превосходит по своим параметрам метод опорных векторов. На практике метод обладает следующими свойствами:

1 Порядка 94% верных обнаружений, при порядка 0,12% ложных срабатываний, для задачи распознавания лиц;

2 Склонность к переобучению не исследована. Следует отметить, что за счет простой классифицирующей функции, можно ожидать хорошей устойчивости к переобучению;

3 Устойчивость к шуму не исследована, но теоретически должна быть достаточно высокой;

4 За счёт просеивания компонент вектора признаков достигается высокая скорость - выше, чем у систем на основе метода опорных векторов.

Метод усиления слабых классификаторов

Усиление слабых классификаторов - это подход к решению задачи классификации, путём комбинирования примитивных классификаторов в один более сильный комитет. Основная идея метода заключается в итеративной минимизации выпуклого функционала ошибки классификации, путем добавления в комитет очередного слабого классификатора.

Для прикладных систем распознавания объектов на изображениях был построен каскад из комитетов слабых классификаторов, работавший по принципу последовательных приближений. Каскадный подход был специально создан для задачи распознавания объектов на изображениях. Его показатели превосходят все остальные системы:

1 95% верных обнаружений, порядка $1 \cdot 10^{-3}$ % ложных обнаружений для комитета, построенного методом AdaBoost.

2 Вычислительная сложность линейна по количеству итераций обучения. Самая быстрая существующая система - скорость в реальном времени.

Сравнение методов обучения по прецеденту

По показателям работы в реальных системах распознавания объектов на изображениях наиболее удачными оказались алгоритмы усиления слабых классификаторов и SNoW. Оба подхода обеспечивают высокую скорость, высокий уровень обнаружений и низкий уровень ошибок второго рода. Сводные данные для различных методов приведены в таблице.

Таблица. Сводные данные систем

Методы	Процент верных обнаружений	Ошибка второго рода
«Наивный» Байес	~91%	~0,2%
Байесовская сеть	~90%	~0,4%
Нейронные сети	~92%	~1,3%
Метод опорных векторов	~72%	~0,6%
SNoW	~94%	~0,12%
Усиление слабых классификаторов	~94%	~0,001%

Достаточно интересные результаты у системы, основанной на «наивном» Байесовском обучении. Единственным серьёзным недостатком системы является сильная зависимость от качества модуля преобразования. Фактически подход требует пересмотра представления изображения объекта интереса в векторе признаков для каждой конкретной задачи.

Нейронные сети и метод опорных векторов достаточно сильно уступают по показателям вышеперечисленным подходам. Главным недостатком нейросетей является низкая скорость и большое число ошибок второго рода. Метод опорных векторов выгодно отличается скоростью и низким уровнем ошибок второго рода, но имеет очень низкий процент верных обнаружений.

Для того чтобы проанализировать получившиеся результаты сравним различные методы выделения объектов в некотором общем виде. Итак, все четыре метода основываются на оптимизации некоторого функционала

ошибки $F(T, w)$, который, в свою очередь, зависит от обучающей выборки и T параметров классификатора w . Так как обучающая выборка фиксирована, параметр T можно опустить и, фактически, задача сводится к оптимизации $F(w)$ по w . Таким образом, решением, готовым классификатором, будет являться

$$w = \arg \min_w F(T, w) \quad (2)$$

Для нейросети w является набором её весов, для SNoW и метода опорных векторов представляет направляющий вектор разделяющей гиперплоскости. Для boosting w является вектором весов для всех возможных простых классификаторов, то есть, если какой-либо простой классификатор не вошёл в итоговый комитет, его вес полагается равным нулю.

Секрет успеха методов, показавших наилучшие результаты, может заключаться в одном их общем свойстве. Дело в том, что SNoW и boosting производят достаточно «разреженные решения», то есть w содержит множество нулей. В то же время, нейронные сети и метод опорных векторов получают достаточно «плотные» решения – большая часть компонентов w не равна нулю. Это может означать, что данные в задачах распознавания имеют особую структуру, наилучшим образом представимую «разреженной» моделью. На практике это означает, что, несмотря на высокую размерность вектора признаков, количество реально значимых компонент в нём не велико. В предположении о наличии шума «разреженные» методы также могут иметь преимущество, так как наиболее зашумлённые компоненты вектора признаков будут иметь нулевой вес и не влиять на решение. Для конкретной задачи размерность может быть сокращена, что и делается неявно «разреженными» методами.

Несмотря на активное развитие данного направления в последнее время, задачи компьютерного зрения еще очень далеки от своего решения. Более того, до сих пор не существует доминирующей и общепринятой парадигмы, в которой бы работало большинство исследователей.

Список литературы: 1. *Y. Amit*. 2D object detection and recognition: models, algorithms, and networks. - Massachusetts Institute of Technology. The MIT Press. -2002 – 325 с. 2. *А. Кручинин*. Распознавание образов для программистов. <http://blog.vidikon.com/> 3. *Paul Viola and Michael J. Jones*. Rapid Object Detection using a Boosted Cascade of Simple Features. IEEE CVPR, 2001. 4. *В. Вежнев, А. Дегтярева*. Обнаружение и локализация лица на изображении. Компьютерная графика и мультимедиа. Выпуск №1(3)/2003. 5. *А. Вежнев*. Методы классификации с обучением по прецедентам в задаче распознавания объектов на изображениях. Лаборатория КГ и Мультимедиа факультета ВМиК, МГУ им. М.В.Ломоносова, Москва, Россия 6. *B. Bhanu, J. Peng*. Adaptive Integrated Image Segmentation and Object Recognition. - IEEE Transactions on systems, man, and cybernetics—part C: Applications and reviews, vol. 30, no. 4, - November 2000 7. *R. E. Schapire, Y. Singer*. Improved Boosting Algorithms Using Confidence-rated Predictions. Machine Learning, 37, 297–336, 1999.

Поступила в редколлегию 07.05.2011